

CNA

AD 643740

THE LAGRANGE MULTIPLIER —
A HEURISTIC PRESENTATION

By W.R. Nunn

OEG Research Contribution No. 76

This research contribution does not necessarily represent the views of CNA or the U.S. Navy. It may be modified or withdrawn at any time.

277

700

CLEARINGHOUSE
FOR FEDERAL SCIENTIFIC AND
TECHNICAL INFORMATION

Hardcopy

Microfiche

\$3.00

\$.65

40

pp

00

1 ARCHIVE COPY

Distribution of this
document is unlimited.

CONTRACT NONR 3732(00)

Research Contribution

DDC
DEC 21 1966
R

OPERATIONS EVALUATION GROUP
Center for Naval Analyses
THE FRANKLIN INSTITUTE
WASHINGTON, D.C. 20350

OEG RESEARCH CONTRIBUTION NO. 76

Operations Evaluation Group

CENTER FOR NAVAL ANALYSES

THE LAGRANGE MULTIPLIER - A HEURISTIC PRESENTATION

By W.R. Nunn

W. R. Nunn

31 October 1966

Work conducted under contract NONR 3732 (00)

Distribution of this
document is unlimited.

ABSTRACT

Many problems in operations research require the maximization or minimization of a suitable payoff function subject to various constraints. Lagrange multipliers are classically used for this type of problem, however, the treatment given this technique by most texts requires that the payoff and constraint functions be at least differentiable at the extremizing point. This paper shows that the Lagrange multiplier concept can be independent of differentiability or even continuity of the functions involved. It also gives the reader a geometric insight into the workings of the multiplier. Simplifications possible, if the functions involved are homogeneous, are displayed. The treatment is heuristic rather than formally rigorous.

TABLE OF CONTENTS

	Page
I. Introduction	1
II. The Direct Method - Main Theorem	2
III. Interpretation of the Lagrange Multiplier	5
A. The g, H plane	5
B. Definition of $\max H, \min H$	7
C. The Lagrange Multiplier as a partial derivative	8
IV. Various Pathologies	11
V. The Cell Problem	14
VI. Comparison of the Direct and Classical Methods.	15
VII. Homogeneous Functions.	17
VIII. Generalizations	21
References	22
Appendix A - Miscellaneous Examples	A-1

I. INTRODUCTION

Many problems in operations research require the maximization or minimization of a suitable payoff function subject to various constraints. Lagrange multipliers (LMs) are classically used for this type of problem, however, the treatment given this technique by most texts requires that the payoff and constraint functions be at least differentiable at the extremizing point. This paper indicates that the LM concept can be independent of differentiability or even continuity of the functions involved. The ideas presented follow closely those of Hugh Everett in reference (a). The reader is given a geometric insight into the workings of the LM. Simplifications which may result, if the functions involved are homogeneous, are displayed. The general treatment is heuristic rather than formally rigorous. Many examples (see appendix) have been worked out.

II. THE DIRECT METHOD - MAIN THEOREM

The LM concept retains utility even when the functions involved are almost completely unrestricted in character. The technique for using the LM in this context will be referred to as the "direct method" in contrast to the "classical method" which involves setting partial derivatives equal to zero.

It is desired to maximize (or minimize) a function $H(x)$ subject to constraints $g_i(x) = C_i$, $i = 1, \dots, m$. Except for being real valued, no restrictions are imposed on these functions. The variable " x " may have any set whatever as domain; this set will be called the "strategy set" and will be designated " S ". The theorem which follows sets the stage for the use of the LM in solving this type of problem.

Theorem: (direct method) Let λ_i , $i = 1, \dots, m$ be fixed, arbitrarily chosen, real numbers. Define $F(x)$ as follows:

$$F(x) = H(x) - \sum \lambda_i g_i(x)$$

Suppose that x' is a point in S which maximizes $F(x)$ in comparison with all other points in S . Then x' maximizes $H(x)$ in comparison with all the points x in S which fulfill the constraints $g_i(x) = g_i(x')$.

Proof: By definition of x' , $F(x') \geq F(x)$ for all x in S . Substituting for $F(x)$:

$$H(x') - \sum \lambda_i g_i(x') \geq H(x) - \sum \lambda_i g_i(x) \text{ for all } x \text{ in } S.$$

If attention is restricted to all x in S which fulfill the constraints $g_i(x) = g_i(x')$ for all values of i , then the summations are equal and can be cancelled from the inequality, leaving $H(x') \geq H(x)$ for these particular x in S , which was to be proved.

The following comments are designed to highlight the important points in the theorem, and to anticipate questions which may have occurred to the reader.

Comment 1: The numbers λ_i , $i = 1, \dots, m$ are the LMs and are fixed numbers which are chosen before the maximization is accomplished. This is in sharp contrast to the classical method in which the LMs are left in the

form of undetermined parameters until the end of the process. This is one of the prices that the direct method must pay because it allows the functions involved to be non-differentiable. In special cases it is possible to retain the LMs as parameters in the direct method also (see section VIII).

Comment 2: $F(x)$ is written in the form $H(x) - \sum \lambda_i g_i(x)$. This is also the way in which LMs are used in the classical method. Other forms of $F(x)$ are possible in both methods (section VIII).

Comment 3: The theorem states "Suppose that x' is a point in S which maximizes $F(x)$. . .". Note that $F(x)$ is to be maximized, specifically. The finding of an inflection point or a relative maximum will not do. However, there is no guarantee that a maximizing point, x' , exists in S . The theorem is applicable only when such a point exists.

Comment 4: By merely reversing the inequalities, the proof of the corresponding theorem for the minimum of $H(x)$ is obtained.

Comment 5: The theorem proves that x' yields the absolute maximum of $H(x)$, as opposed to a relative maximum.

Comment 6: The point x' is a point in S which maximizes $F(x)$. In case there are several such points, the proof still holds.

Comment 7: The point x' is obtained by maximizing $F(x)$ without regard to the values taken by the constraint functions. This is precisely why the LMs are valuable; they turn a maximization problem with constraints into a maximization problem without constraints.

Comment 8: The theorem says that x' maximizes $H(x)$ subject to the constraints $g_i(x) = g_i(x')$. The original problem was to maximize $H(x)$ subject to $g_i(x) = C_i$. There is no assurance that $g_i(x') = C_i$ for any or all values of i , hence the original problem has not been solved yet. Recall that the LMs were chosen arbitrarily. It may be possible to cause $g_i(x') = C_i$ for all values of i by suitably modifying the LMs and repeating the process. If this is possible, the direct method solves the original problem. In many OR problems, however, one is not interested in maximizing $H(x)$ subject to specific constraints, but rather in exploring the entire range of what can

be obtained as a function of the constraint values, C_i . In this case the process of sweeping through a range of values of the LMs may be completely adequate.

The direct method, then, boils down to the following sequence:

Step 1: Pick the LMs arbitrarily. Section III will provide guidance for correcting this initial guess, if necessary.

Step 2: Form $F(x)$ and maximize it (or minimize, if desired), obtaining a maximizing point x' , and corresponding values $g_i(x')$ of the constraint function.

Step 3: If $g_i(x') = C_i$ for all values of $i = 1, \dots, m$ then the original problem is solved and the maximum value of $H(x)$, subject to $g_i(x) = C_i$, is given by $H(x')$. If $g_i(x') \neq C_i$ for all values of i , it is necessary to pick new values of the LMs and repeat the process, aiming for convergence of the values $g_i(x')$ to C_i .

An example of the direct method is given in the appendix (example 1). $H(x)$ and $g(x)$ are differentiable, but S is a collection of discrete points, and the classical method does not work.

III. INTERPRETATION OF THE LAGRANGE MULTIPLIER

A. The g,H plane

This section is intended to promote geometric insight as to why the LM works. One constraint function, $g(x)$, will be used in this discussion, but the results hold for any number of constraints. Letting x' denote a point in the set, S , which maximizes $H(x) - \lambda g(x)$ in comparison with all other points in S , we have $F(x') = H(x') - \lambda g(x') \geq H(x) - \lambda g(x)$ for all x in S . Solving this for $H(x)$:

$$H(x) \leq g(x) + F(x') \text{ for all } x \text{ in } S$$

From this basic inequality, most of the results of this section will follow immediately. It will be useful to look at what is happening in the g,H plane.

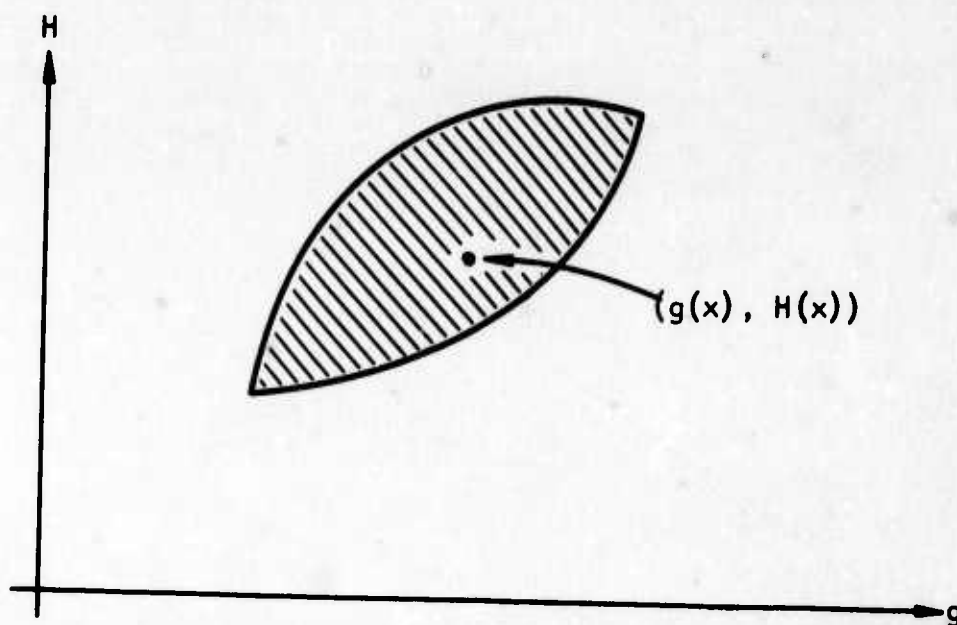


FIGURE 1

As depicted in figure 1, to each point in S there corresponds a point $(g(x), H(x))$ in the g, H plane. In this manner the entire strategy set, S , is mapped into some set, called the payoff set, in the g, H plane. This payoff set is shown as the shaded region in figure 1. In general there is no way of predicting just what this payoff set is going to look like.

The inequality $H(x) \leq \lambda g(x) + F(x')$ for all x in S says that all the points in the payoff set lie below the line whose equation is $H = \lambda g + F(x')$. Furthermore, it says that this line touches the payoff set only at the maximizing point (or points, in case of non-uniquity). This immediately suggests that the act of guessing a value for λ in step 1 of the direct method is no more than guessing the slope of a line in the g, H plane. The act of maximizing $F(x)$, step 2, to find $F(x)$, simply pushes the H -intercept up the H axis until the line $H = \lambda g + F(x')$ just touches the payoff set at the point $(g(x'), H(x'))$ as shown in figure 2. In brief, λ is the slope of the line, and the $F(x')$ is the H -intercept.

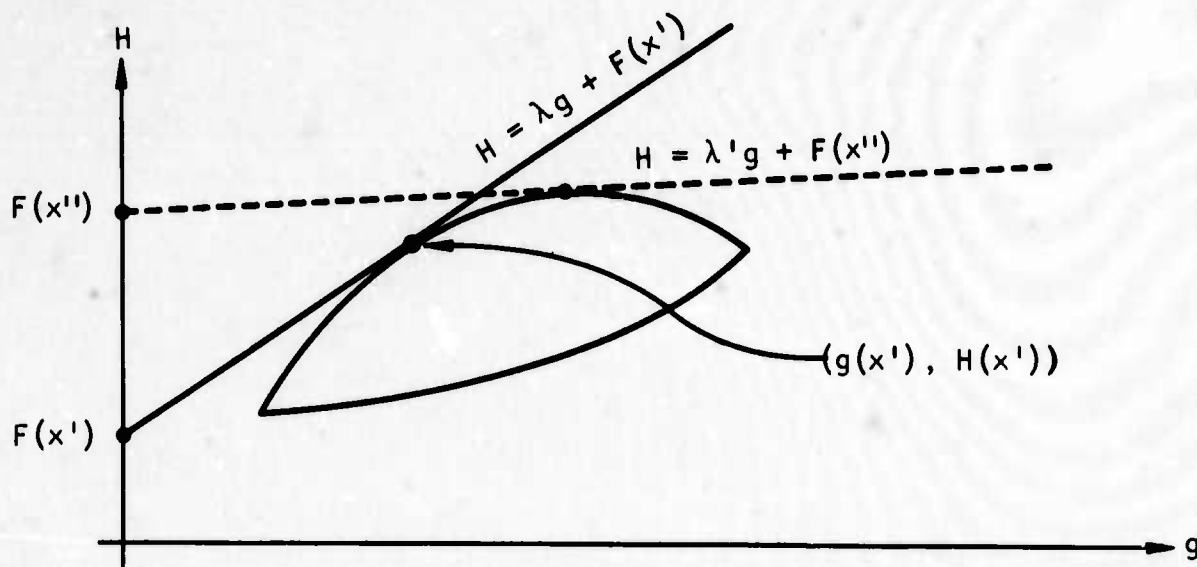


FIGURE 2

Figure 2 also illustrates how guessing new values of λ produces new values of the constraint function $g(x)$. If the value $g(x')$, initially obtained by steps 1, 2, is not large enough, then λ must be made smaller, i.e. the slope of the line must be decreased. The resulting line is illustrated in the figure above as $H = \lambda'g + F(x'')$. It is obvious that the corresponding value $g(x'')$ cannot possibly be less than $g(x')$. This basic concept will help the analyst modify his guesses of λ in the proper direction to obtain the desired constraint values.

Figure 2 illustrates another, very important point. Suppose that after choosing a value for the LM we obtain a maximizing point x' and a corresponding value $g(x')$ of the constraint. We know from the main theorem that this x' maximizes $H(x)$ subject to the constraint $g(x) = g(x')$. Suppose further that the value of the LM chosen was positive. Then figure 2 illustrates that you cannot do better, i.e., obtain a larger $H(x)$, by using any smaller value of constraint than $g(x')$. This follows directly from the basic inequality obtained at the beginning of this section, and is one of the main advantages of using the linear form of $F(x)$. Again, if the LM is positive, then x' maximizes $H(x)$ subject not only to $g(x) = g(x')$, but also to $g(x) \leq g(x')$. This is of paramount importance in problems where $g(x)$ is a money or other resource constraint, and $H(x)$ is the payoff for expending $g(x)$ units of the resource.

B. Definition of maxH, minH

After the maximization operation is completed successfully, i.e., a point x' is found which maximizes $H(x)$ subject to $g(x) = C$, it is intuitively clear that x' depends upon the value of C . (Otherwise, the constraint is having no effect on the problem and may as well be dispensed with.) It follows that $H(x')$, the resulting maximum value of $H(x)$, is also a function only of C . This function is denoted $\max H(C)$, and the corresponding function resulting from minimizing will be denoted $\min H(C)$. Example 2 displays a typical $H(x)$ and the resulting $\min H(C)$.

These functions have an important interpretation in the g, H plane. The line in the g, H plane whose equation is $g = C$ is assumed to intersect the payoff set as shown below.

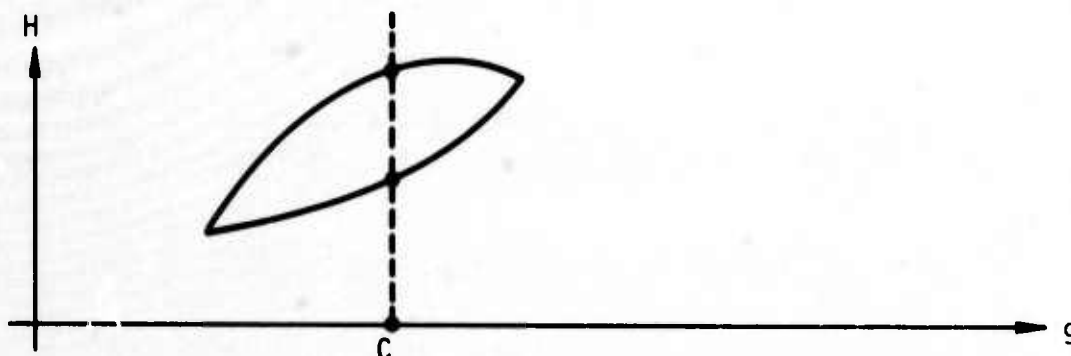


FIGURE 3

The points in the payoff set which fall on this line are precisely those for which $g(x) = C$. The point where the line $g = C$ cuts the upper boundary of the payoff set is the desired maximum value of $H(x)$ subject to the constraint $g(x) = C$. It follows, therefore, that the entire upper boundary of the payoff set is the graph of the function $\max H(C)$ (the lower boundary is the graph of $\min H(C)$) provided only that the axes in figure 3 are relabeled $\max H(C)$, C instead of H, g respectively. It is emphasized that $H(x)$ and $\max H(C)$ are completely different functions. The properties of these two functions may be strikingly different, as will be illustrated in the following subsection.

C. The Lagrange Multiplier as a partial derivative

Figure 4 displays the typical situation resulting from the successful use of the direct method. For simplicity of notation and graphing only one constraint is used.

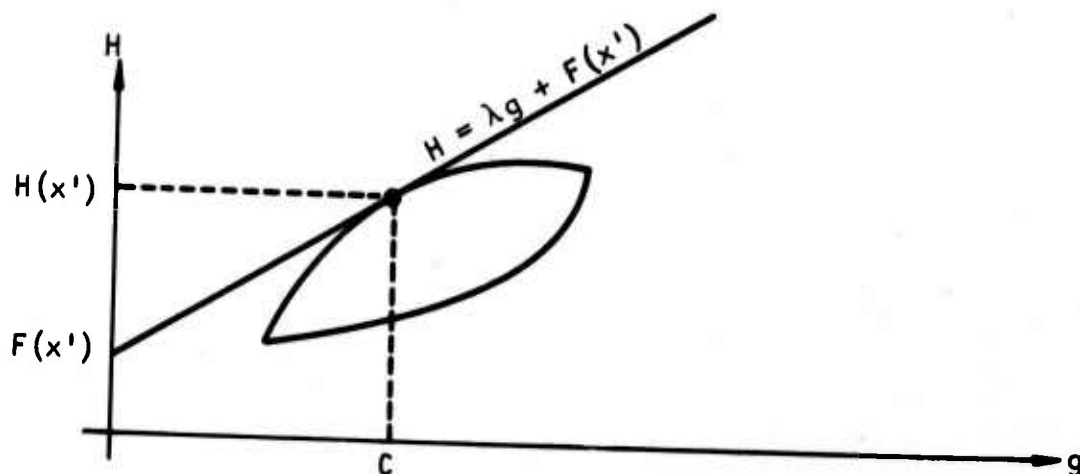


FIGURE 4

A maximizing point x' has been found which maximizes $F(x)$. Also $g(x') = C$, the desired value. Assume that the function $\max H$ is differentiable at this value C . If so, the line $H = \lambda g + F(x')$ and the function $\max H$ must have the same slope at the point where the line touches the payoff set. The slope of the line is λ and the slope of $\max H$ at this point is $d(\max H)/dc$ (partial derivative in the case of multiple constraints). This is the desired interpretation of the LM. The LM is the derivative of $\max H(C)$ with respect to C .

This interpretation of the LM will frequently permit a ballpark guess as to the proper value of the LM when using the direct method. This approach is particularly useful when the functions $\max H(C)$, $H(x)$, and $g_i(x)$ have physical or geometric interpretations (example 3).

Algebraic constraints may usually be formulated in a number of equivalent ways. For example, $g(x, y) = ax^2 + bxy + cy^2 = D$ can be rewritten as $(ax^2 + cy^2 - D)/xy = B$. In the first case the LM will be $d(\max H)/dD$, in the second, $d(\max H)/dB$. Since the effect of varying D is to expand or contract the geometric figure $g(x, y) = D$ while varying B rotates the geometric figure, the numerical value of the LM will be different in the two cases. As a result, an apparently trivial reformulation of a constraint may introduce or remove a pathology from the problem.

The foregoing interpretation of the LM as a partial derivative depends upon $\max H(C)$ being differentiable. Frequently this is not the case; this can occur in various ways. One of the simplest is for the strategy set S to have only a finite number of points, as in example 1. $H(x)$ is differentiable throughout most of the region of interest, but the payoff set contains only five points, hence $\max H(C)$ is nowhere differentiable. On the other hand, it is possible for $\max H(C)$ to be differentiable even when some or all of the functions $H(x)$ and $g_i(x)$ are non-differentiable. In example 8, $g(x)$ is discontinuous at the minimizing point for all values of C greater than -1.90 , yet $d(\min H)/dC$ exists and equals λ for all such values of C .

IV. VARIOUS PATHOLOGIES

Frequently there are several values of x in S which will maximize $F(x)$ for a particular value of the LM. One common cause of this is illustrated below.

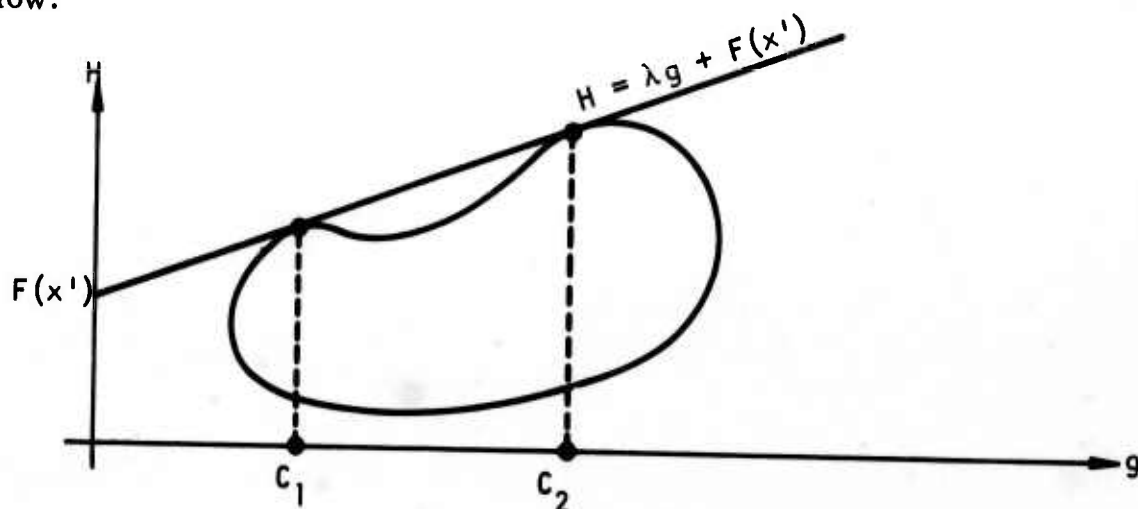


FIGURE 5

The payoff has a concavity in the upper envelope, allowing the line $H = \lambda g + F(x')$ to be tangent at two places. This illustrates another point. It is impossible to find a LM which will give rise to values between C_1 and C_2 without violating the inequality $H(x) \leq \lambda g(x) + F(x')$, which was obtained at the beginning of section III. Such regions are called "gaps", characterized by a small change in the LM causing a large jump in the resulting value of the constraint. One way of overcoming the effect of a gap will be discussed in section VIII; another method is mentioned in reference (a).

In figure 5, above, two values of x in S are obtained which maximize $F(x)$; one gives the value C_1 , the other, C_2 . It is possible for there to be an infinite number of points in S which maximize $F(x)$ for a certain value of the LM. The most common cause for this phenomenon is illustrated in figure 6, next page.

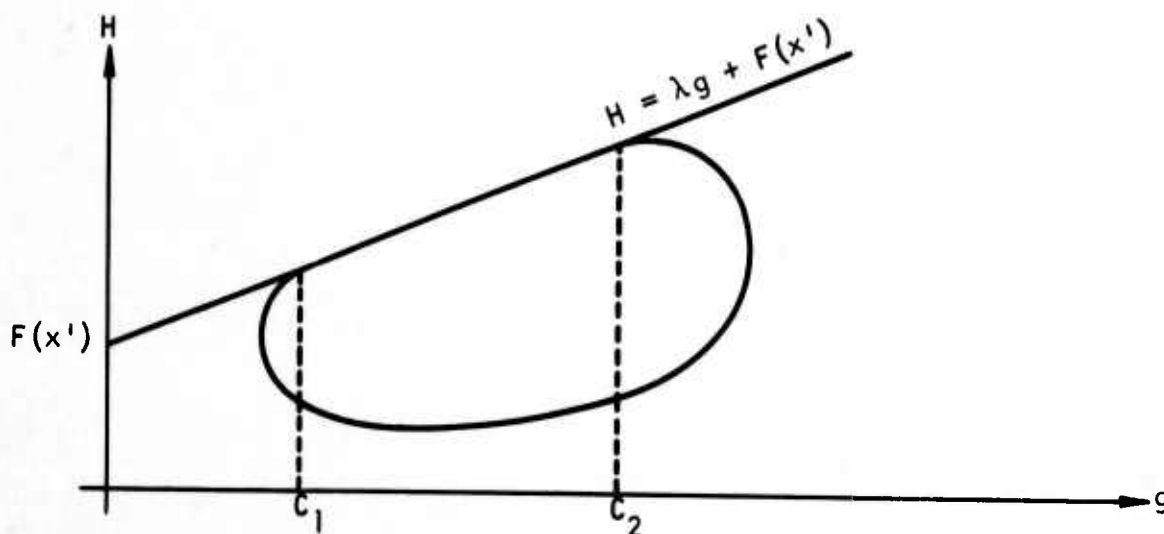


FIGURE 6

Between C_1 and C_2 the function $\max H(C)$ is a straight line segment with slope λ . All of the points x' in S which map onto this line segment will maximize $F(x)$ for this particular value of the LM, producing the same $F(x')$. This phenomenon, the "constant multiplier", will probably be fatal if a computer is being used to solve the problem, unless the program is carefully written. Unless exactly the correct value of the LM is guessed, it will not be possible for the computer to arrive at any value of C between C_1 and C_2 . On the other hand, if the exact value of the LM is guessed, the computer may try to print out all the points in S which maximize $F(x)$ for this value of the LM. Hence some care must be taken in the programming of the direct method to test for non-unicity of x' before ordering a printout (example 4).

The preceding figures illustrate that one value of the LM may give rise to many maximizing points x' . The opposite situation may occur, many values of the LM may give rise to the same x' . This situation is very common, arising primarily by $\max H(C)$ not being differentiable. Figure 7, next page, illustrates how this may occur.

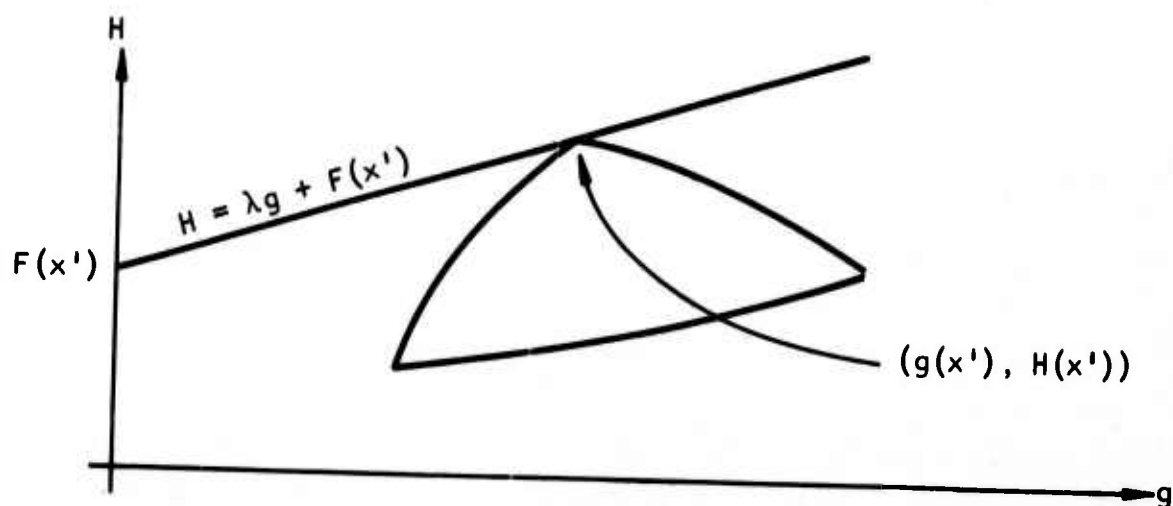


FIGURE 7

The function $\max H$ has a sharp corner at one point, i.e. $\max H$ is not differentiable there. It is clear that there are many slopes such that the resulting line will touch the payoff set, as shown in figure 7. All these values of the LM will give the same value for $g(x')$ and for $H(x')$.

V. THE CELL PROBLEM

There is a type of problem, the cell problem, for which the LM technique, direct or classical, is particularly well suited. It is characteristic of this problem that $F(x)$ can be written in the form

$$\sum H_i(x_i) - \lambda \sum g_i(x_i)$$

In other words, in the cell problem we have $H(x) = \sum H_i(x_i)$ and $g(x) = \sum g_i(x_i)$. For simplicity of notation only one such constraint is shown in $F(x)$, but any number, each with its own LM, is permitted, as usual. It is clear that $F(x)$ can be rewritten as a single sum

$$\sum (H_i(x_i) - \lambda g_i(x_i))$$

of "n" terms, hence the name. It follows that $F(x)$ can be maximized by maximizing each cell separately (using the same value of LM), and summing the results. A computer is especially useful for cell problems involving a large number of cells because computing time increases linearly with the number of cells rather than exponentially. Even if the maximization in each cell is done analytically the computer is useful in summing results for each value of LM, and in otherwise "keeping the books".

Typical problems of this type are weapon allocation or weapon mix problems. Example 4 is of this type. Reference (a) discusses cell problems in some detail. Reference (b) contains an example of a cell problem in which the casualty rates in several shipping channels due to mine laying is maximized subject to constraints on the number and types of mines laid.

VI. COMPARISON OF THE DIRECT AND CLASSICAL METHODS

It is desired to maximize via the classical method a function $H(x)$ subject to constraints $g_j(x) = C_j$, $j = 1, 2, \dots, m$. Here, " x " is a point in Euclidean " n " space, i.e., " x " denotes (x_1, \dots, x_n) . The notation " C " and " λ " will similarly be used to denote the " m " space vectors (C_1, \dots, C_m) and $(\lambda_1, \dots, \lambda_m)$. This is in contrast to the direct method where there are no restrictions on the character of the argument of $H(x)$. As normally presented, the classical method consists of the following steps:

Step 1: Form $F(x) = H(x) - \sum \lambda_j g_j(x)$ where the LMs are parameters rather than specific values.

Step 2: Write down the system of " n " equations

$$F_{x_i} = 0$$

where F_{x_i} denotes the partial derivative of F with respect to x_i , evaluated at " x ".

Step 3: These equations plus the constraint equations provide $m+n$ equations in the $m+n$ unknowns, $x_1, \dots, x_n, \lambda_1, \dots, \lambda_m$. Solve these equations for the unknowns. The resulting solutions (there may be several) represent stationary points for $H(x)$ which must be tested separately to determine which represent relative extrema, stationary points only, or absolute extrema. Example 5 illustrates this method.

The following comments are intended to clarify classical methods and to help pinpoint the differences between the two methods:

Comment 1: In step 1, the LMs are parameters rather than specific numbers. In order for this to work, the maximizing point x' must be a smooth, continuous function of both C and λ . This is because step 3 is equivalent to solving for the maximizing vector x' as this function of C and λ , then using the constraint equations to eliminate λ . When non-differentiable functions are used, the function $x'(C, \lambda)$ may not exist, or be expressible in any reasonable form.

Comment 2: Step 2 is analagous to maximizing or minimizing a function of one variable by setting its derivative equal to zero. In the event that this procedure locates the absolute maximum or minimum of $F(x)$ then the main theorem of section I provides a proof for the method.

Comment 3: Since the absolute maximum of $F(x)$ is not the goal of steps 2, 3, the inequality of section III does not hold in the classical method. This has two immediate consequences:

(a) The pathologies of section IV do not occur. In the case where a gap is present, the classical method will find the desired x' which maximizes $H(x)$, but this x' does not maximize $F(x)$ at the same time. $F(x)$ does have a stationary point at x' , however. Example 6 illustrates a problem which is solved immediately by the classical method, but not by the direct method unless modifications are introduced to eliminate the gap.

(b) The line $H = \lambda g + F(x')$ may pass through the payoff set without being anywhere tangent. Such solutions correspond to the relative extrema which are not absolute extrema, or to stationary points which are neither.

Comment 4: When the classical method obtains the absolute extrema the resulting LMs can be interpreted as partial derivatives of $\max H$ or $\min H$ in analogy to the direct method. This is suggested by the following: Multiply each of the equations $F_{x_i} = 0$ by d_{x_i} and sum, obtaining $dF(x) = 0$, and hence $dH(x) = \sum \lambda_j dg_j(x)$. Evaluating this latter equation at the extremizing point x' obtained in step 3, we have

$$dH(x') = \sum \lambda_j dg_j(x').$$

Intuitively, x' is a function of C , denoted $x'(C)$. $H(x')$ is therefore $H(x'(C))$, which is, in fact, $\max H(C)$ or $\min H(C)$ since x' was extremizing. This, plus the equations $g_j(x') = C_j$, suggests that $d(\max H(C)) = \sum \lambda_j dC_j$ (or $\min H$, as desired) and hence that λ_j is the partial derivative of $\max H(C)$ (or $\min H(C)$) with respect to C_j , as desired.

VII. HOMOGENEOUS FUNCTIONS

When $H(x)$ and $g_i(x)$ are homogeneous functions, a good deal is known about the resulting functions $\max H$ and $\min H$. In particular, the single constraint problem becomes quite susceptible to attack by direct method. A function $f(x)$ is homogeneous, of degree "p", if $f(ax) = a^p f(x)$ for all values of "a" (for this discussion, "x" will be treated as having components " x_i " but no restrictions on the type of number of these components is desired). The function $f(x, y) = x^3 + xy^2$ is homogeneous of degree 3. The function $f(x, y, z) = \max(x, y, z^2/y)$ is homogeneous of degree 1, and illustrates a discontinuous homogeneous function. The homogeneity property imposes a certain regularity of behavior upon the function, nevertheless, homogeneous functions may be exceedingly complex or otherwise intractable. Fortunately, homogeneity can frequently be determined by inspection in spite of the complexity.

An "isobaric" function is a useful generalization of a homogeneous function. A function $f(x)$ is isobaric if a constant p and a vector q exist such that $f(a^q x) = a^p f(x)$ for all values of a , where the notation " $a^q x$ " denotes the vector with components " $a^{q_i} x_i$ ". The function $f(x, y) = |xy^2|$ is isobaric since $|(a^2 x)(ay)^2| = a^4 |xy^2|$, i.e., $p, q_1, q_2 = 4, 2, 1$, respectively. It is easily seen that if all values q_i are equal, the isobaric function reduces to a homogeneous function of degree p/q_i .

Theorem 1: Let $H(x)$ be homogeneous of degree p , and let each of the functions $g_i(x)$ be homogeneous of degree q_i . If the function $\max H$ (or $\min H$) exists, it is isobaric, and $\max H(a^q C) = a^p \max H(C)$.

proof: Let S_1 be the subset of S for which the vector equation $g(x) = C$ holds. Let S_2 be the subset of S for which $g(x) = a^q C$. Using the homogeneity of each function g_i , it follows that if x is a point in S_1 then ax is a point in S_2 (multiply both sides of the equation $g_i(x) = C_i$ by a^{q_i}). Conversely, any point z in S_2 can be written as ax where x is a point in S_1 . It follows therefore, that

$$\max H(a^q C) = \max_{z \in S_2} H(z) = \max_{x \in S_1} H(ax) = \max_{x \in S_1} a^p H(x) = a^p \max H(C).$$

Theorem 2: If $f(x)$ is isobaric, then $f(x)$ has a derivative, denoted $f'(x)$, along the curve $a^q x$, where a is a parameter. Furthermore, $f'(x) = (p/qx)f(x)$.

proof: $f'(x) = \lim_{a^q \rightarrow 1} \frac{f(a^q x) - f(x)}{a^q x - x} = \lim_{a^q \rightarrow 1} \frac{a^p f(x) - f(x)}{a^q x - x} = \frac{f(x)}{x} \lim_{a^q \rightarrow 1} \frac{a^p - 1}{a^q - 1} = (p/qx)f(x).$

Corollary 1: Under the conditions of theorem 1, if there is only one constraint then $\max H(C)$ and $\min H(C)$ (if they exist) are of the form $AC^{p/q}$, where A is a constant.

proof: Using theorems 1 and 2, the functions $\max H$ and $\min H$ satisfy the differential equation

$$df(C)/dC = (p/qC)f(C)$$

which has the general solution $f(C) = AC^{p/q}$.

Comment 1: This corollary is the primary result of this section. If the direct method is used, only one guess for the value of the LM need be made since the resulting values of C and $\max H(C)$ can be used to evaluate "A" and obtain the entire solution. This finds application in the classical method, also. After the partial derivatives have been equated to zero, the resulting system may be impossible to solve in terms of the LM as a general parameter. It suffices, however, to solve the system for only one value of the LM. Example 7 illustrates these ideas.

Comment 2: A gap or constant multiplier can be immediately spotted by examining p/q . A simple technique for remedying an unfavorable p/q will be presented in section VIII.

Comment 3: The proof of theorem 2 does not require the isobaric property for all values of a , merely for values of a in some neighborhood of $a = 1$. Similarly, theorem 1 remains valid for all values of a for which the functions

involved are homogeneous. Hence, corollary 1 holds providing only that the functions involved are all homogeneous for values of a within some neighborhood of $a = 1$. The utility of this is that many functions occurring in practical problems are homogeneous for restricted values of a , e.g., functions involving absolute values may require a to be non-negative.

Comment 4: Functions which are piecewise defined as, for example,

$$f(x, y) = \begin{cases} x + 3y, & x \geq 2 \\ 2x + y, & x < 2 \end{cases}$$

are not necessarily homogeneous along the boundaries common to the various pieces. This is in spite of the fact that the degree of homogeneity may be the same within each separate piece. The problem is that of proper continuity at the boundary (see comment 3, above). Homogeneous functions may be discontinuous, but their discontinuities must be of a rather specific type; in particular, the homogeneous function must be continuous along all radial curves.

There is an analogous result for the multi-constraint case.

Lemma 1: If the function $\max H$ is differentiable, it obeys the first order linear partial differential equation

$$\sum q_i x_i f_{x_i} = p f$$

proof: Differentiate both sides of the equation $f(a^q x) = a^p f(x)$ with respect to a , and let a go to 1.

Comment 1: The differentiability of the function f must be assumed. Theorem 2 guarantees the existence of only one directional derivative; the partials may not exist.

Comment 2: This partial differential equation is readily solved, giving

$$G(f^q/x^p) = 0$$

where G is an arbitrary function. If a suitably well behaved boundary condition were available in the form $\max H(C) = M(t)$ along a known curve $C(t)$, where t is a parameter, then the arbitrary function G could be analytically determined. In most practical applications, the boundary conditions will be analytically

unsuitable, or unknown. In the former case, numerical integration of the partial differential equation may be successful; in the latter case, the direct method may be tried. It may be possible to approximate a boundary condition arbitrarily closely by varying the LMs so as to sketch out $M(t)$ along some arbitrary curve $C(t)$.

VIII. GENERALIZATIONS

In applications of the classical method, it is common to maximize $\log(H(x))$ or $(H(x))^2$ if these are more tractable analytically, since "maximizing these is the same as maximizing $H(x)$ ". In general it is true that if G is any monotonically increasing (strictly) function, then $G(H(x))$ attains its maximum at the same point x' as the function $H(x)$. This fact is of fundamental importance in the direct method because the function G changes the shape of the payoff set. A proper choice of G can eliminate a gap or can eliminate the "constant multiplier" without changing x' (example 6).

The constraint functions may also be modified into $G(g(x)) = C$. This modification also changes the shape of the payoff set and can be used instead of, or in addition to, modifying $H(x)$. These two types of modifications have inverse effects on the shape of the payoff set, however (example 6).

The single constraint problem with homogeneous functions can often be effectively handled using these concepts. If the ratio of degrees of homogeneity (p/q of the preceding section) is unfavorable, G can be chosen homogeneous of the proper degree to guarantee that no gap exists (example 6).

It is sometimes possible to retain the LMs as parameters in the direct method and still perform the maximization required in step 2, section II. This requires that x' be obtainable as a function of the parameter λ . If non-differentiable functions are involved, then $x'(\lambda)$ will probably be non-analytic and must be examined point at a time. Example 8 illustrates a case where the classical method will not work, but the direct method applies; the maximization (minimization in this case) can be done analytically, retaining the LM throughout as a parameter.

References: (a) "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources", by H. Everett III. ORSA Journal, Vol. 11, No.3, May - June, 1963

(b) OEG Study 639, Classified.

APPENDIX A

MISCELLANEOUS EXAMPLES

Example 1: This problem illustrates the direct method when the set S is a collection of discrete points.

Maximize $|x^2 - 2y|$ subject to $x + y^2 = 3$ where S consists of the collection of points $(0, 1)$, $(2, 1)$, $(4, 1)$, $(3, 0)$, $(1, 2)$.

Step 1: $F(x, y) = |x^2 - 2y| - \lambda(x + y^2)$.

Step 2: It is apparent that in order to maximize $F(x, y)$ we must plug in each of the points in S in turn and evaluate $F(x, y)$ at each point. Further, we cannot investigate $F(x, y)$ numerically for a maximum so long as the LM is a parameter. It is necessary to insert a numerical value for λ , then proceed with that value into step 3. Listed below are the results of three guesses for λ .

λ	Maximizing point (x')	$H(x')$	$g(x')$	$F(x')$
10	(0, 1)	2	1	-8
1	(4, 1)	15	5	9
3	(3, 0)	9	3	0

The first guess, $\lambda = 10$, gave a maximum of -8 to $F(x, y)$, the maximizing point being $(0, 1)$. Proceeding into step 3, $g(0, 1)$ turns out to be 1. Since the original problem specifies the constraint $g(x, y) = 3$, the value of 10 for the LM is unsatisfactory and another guess must be made. In section II it will be made clear that the LM must be decreased in order to increase the resulting value of $g(x, y)$. A guess of 1 for λ produces a constraint value of 5, i.e., the LM has been decreased too much. A guess of 3 for λ finally produces $g(x, y) = 3$, as desired. According to the main theorem, section I, the point $(3, 0)$ maximizes $H(x, y)$ subject to $g(x, y) = 3$.

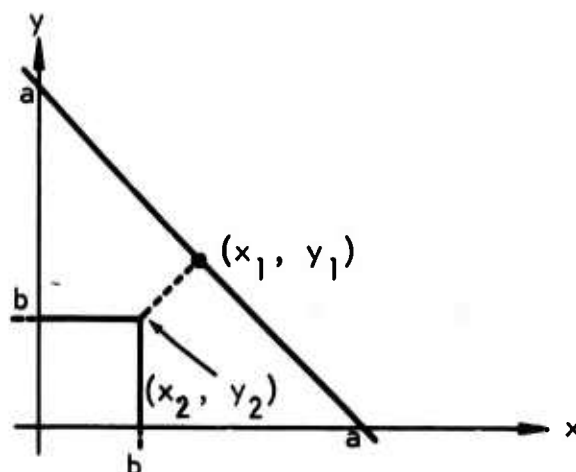
Similar problems involving larger sets S , more constraints, and more complex functions can be solved in exactly this way, using a computer to evaluate the functions.

Example 2: This problem illustrates how the function "minH" comes into being.

It is desired to find the minimum distance from the origin of the x, y plane to the curve defined by $xy = b$, for $b > 0$. The problem is trivial and can be solved readily in a variety of ways; by sketching the graph of $xy = b$ for some value of b it will become obvious that the solution must be $x = y = (b)^{1/2}$. For this problem $H(x, y) = (x^2 + y^2)^{1/2}$. Plugging in the solution, $H(x, y)$ becomes $(b + b)^{1/2}$ or $(2b)^{1/2}$, which is no longer a function of x and y , but of the parameter b . This resulting function of b is $\text{minH}(b)$.

Example 3: This problem illustrates how the geometric interpretations of the functions $\min H(C)$ and $g(x)$ may be used to estimate the values of the LMs. In this example, these considerations permit an exact solution.

It is desired to find the minimum distance between the line $x + y = a$ and the square $\max(|x|, |y|) = b$, where the values of a and b are such that the two figures are not intersecting. In this problem, $H(x_1, y_1, x_2, y_2) = ((y_1 - y_2)^2 + (x_1 - x_2)^2)^{1/2}$, $g_1(x_1, y_1, x_2, y_2) = x_1 + y_1$, and $g_2(x_1, y_1, x_2, y_2) = \max(|x_2|, |y_2|)$, where the subscript "1" refers to the line, and "2" refers to the square. The reader is requested to examine the representative figure below and use the interpretation of the LMs as partial derivatives, along with the geometric information, to evaluate the LMs directly.

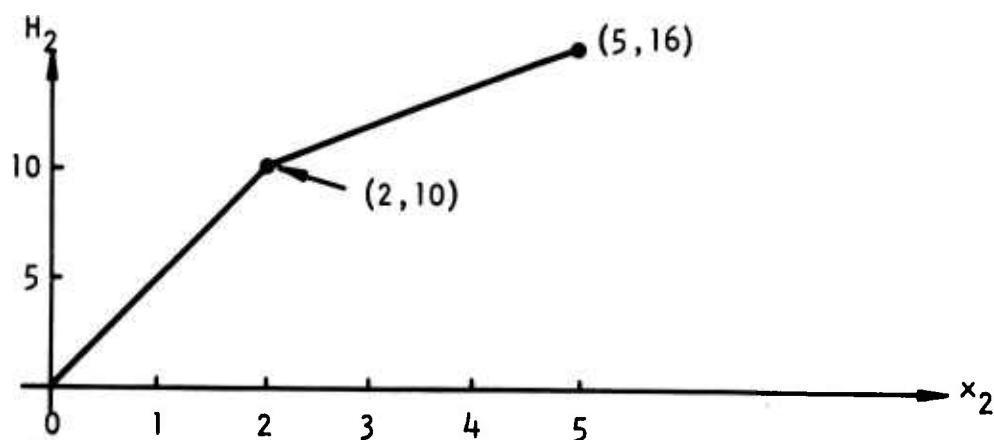


Solution: The function $\min H(a, b)$ has the geometric interpretation of minimum distance, hence λ_1 is the derivative of this distance with respect to "a", and λ_2 is the derivative with respect to "b". An increase of a by one unit causes an increase in $\min H(a, b)$ by $1/(2)^{1/2}$ units, hence $\lambda_1 = 1/(2)^{1/2}$. Similarly, an increase of b by one unit causes a decrease in $\min H(a, b)$ by one unit, hence $\lambda_2 = -1$.

Note that the values of the partial derivatives are independent of both a and b . This is untypical; usually the LMs are functions of all of the constants.

Example 4: This is a cell problem in which $H(x)$ is non-differentiable.

Maximize $H(x) = H_1(x_1) + H_2(x_2)$ subject to the constraint $g(x_1, x_2) = x_1 + x_2 = 5$, where $H_1(x_1) = x_1(10 - x_1)$ and $H_2(x_2)$ is as shown.



$F(x)$ can be written in the form

$$\left[H_1(x_1) - \lambda x_1 \right] + \left[H_2(x_2) - \lambda x_2 \right]$$

and each of the terms in brackets can be maximized separately. The first bracket can be maximized analytically, yielding a maximum at $x_1 = 5 - \lambda/2$. The second bracket is not differentiable at all points and must be maximized by considering separate values of λ . This can be done by inspection, in this example, since the graph of $H_2(x_2)$ is available; however, if H_2 were more complex, it might be necessary to divide up the x_2 axis into small increments and let a computer search for the maximum. The table below summarizes the results of various guesses of the LM.

λ	x_1	x_2	$x_1 + x_2$	$H_1(x_1, x_2)$	$H_2(x_1, x_2)$	$H(x_1, x_2)$	$F(x_1, x_2)$
2	4	2-5	6-9	24	10-16	34-40	22
3	3.5	2	5.5	22.75	10	32.75	16.25
4	3	2	5	21	10	31	11

The point $x' = (3, 2)$ satisfies the constraint $g(x_1, x_2) = x_1 + x_2 = 5$, hence it is the desired point, giving a maximum value of 31 to $H(x_1, x_2)$. Note that for $\lambda = 2$, the maximizing point x' is not unique since x_2 may vary between 2 and 5, producing a corresponding variation in $H(x_1, x_2)$ (but not in $F(x_1, x_2)$). This is the "constant multiplier" discussed in section IV. Note that as the LM is varied from 3 to 4, x_2 remains constant at 2. This is because H_2 is non-differentiable, and is discussed in section IV.

As the LM is varied from 2 to 4, various values of $g(x_1, x_2)$ and their accompanying values of $H(x_1, x_2)$ are also generated. The plot of $H(x_1, x_2)$ vs $g(x_1, x_2)$ is discussed in detail in sections III and IV, and is fundamental to the study of LMs.

This problem is handled in a straightforward way using the direct method. The classical method cannot handle it since $H_2(x_2)$ is not differentiable at the maximizing point.

Example 5: This problem illustrates a straightforward application of the classical method.

Find the minimum distance from the origin to the subspace determined by the intersection of the two hyperplanes

$$a \cdot x = C_1$$

$$b \cdot x = C_2$$

where the dot product notation has the usual interpretation

$$a \cdot x = \sum_{i=1}^n a_i x_i$$

Solution: For this problem, $H(x) = (x \cdot x)^{1/2}$, however the square of this will be minimized instead since it is more tractable analytically. This technique is discussed in section VIII.

Step 1: $F(x) = x \cdot x - \lambda_1 a \cdot x - \lambda_2 b \cdot x$

Step 2: Equating the partial derivatives to zero gives the vector equation

$$2x - \lambda_1 a - \lambda_2 b = 0$$

from which x can immediately be expressed as a function of the LMs.

$$x = 1/2 (\lambda_1 a + \lambda_2 b)$$

Step 3: This vector x can be plugged into the constraint equations, giving:

$$\lambda_1 a \cdot a + \lambda_2 a \cdot b = 2C_1$$

$$\lambda_1 b \cdot a + \lambda_2 b \cdot b = 2C_2$$

Since the dot products can be calculated directly, the above equation in the two unknowns λ_1, λ_2 can be solved by Cramer's rule to give the LMs as functions of C_1, C_2 . The determinant D is

$$(a \cdot a) (b \cdot b) - (a \cdot b)^2$$

which, by the Cauchy-Schwarz inequality, is non-zero unless the vectors a and b are colinear. In this case, the two planes are parallel and intuition agrees that there is no solution. Solving for the LMs

$$\lambda_1 = 2(C_1 b \cdot b - C_2 a \cdot b)/D$$

$$\lambda_2 = 2(-C_1 a \cdot b + C_2 a \cdot a)/D$$

These values can be plugged into the expression for x , giving x as a function of C_1, C_2 . The expression $x \cdot x$ can then be evaluated giving, finally, $\min H(C_1, C_2)$.

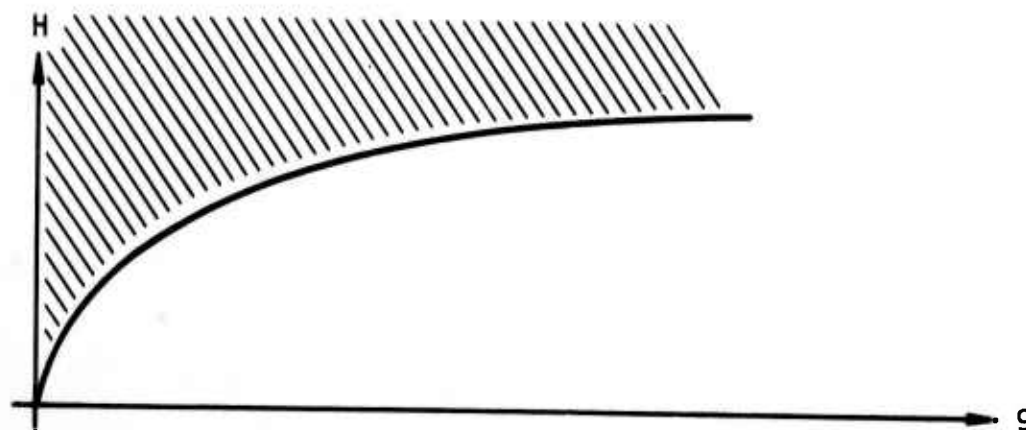
A faster method, however, is to take the equation for x , dot both sides with " x ", and evaluate at the minimizing point, giving:

$$\begin{aligned} \min H(C_1, C_2) &= 1/2(\lambda_1 C_1 + \lambda_2 C_2) \\ &= \frac{C_1^2 b \cdot b - 2C_1 C_2 a \cdot b + C_2^2 a \cdot a}{D} \end{aligned}$$

This savings in labor is always possible if the functions $H(x)$ and $g_i(x)$ are homogeneous (section VIII).

Example 6: This problem can be solved immediately by the classical method, but cannot be solved by the direct method unless one or more of the functions is modified to eliminate the "gap". A few possible modifications are illustrated.

Minimize $(x^2 + y^2)^{1/2}$ subject to the constraint $g(x, y) = xy = b$, with $b > 0$. S is the first quadrant of the x, y plane. The classical method gives the solution $x = y = (b)^{1/2}$, $\lambda = 1/(2b)^{1/2}$, $\min H(b) = (2b)^{1/2}$. The payoff set in the g, H plane is shown below.



The function $\min H(b)$ is concave downward, hence the entire positive g axis (except the origin) is a "gap" and the direct method will fail (the inequality of section III is reversed here since a minimum is desired).

We can modify the shape of the payoff set so as to eliminate the gap by modifying either $H(x, y)$, $g(x, y)$ or both, appropriately. If $G(H)$ is a monotonically increasing function of H in the region of interest, then $G(H(x, y))$ can be used in place of $H(x, y)$. Furthermore, the gap will be reduced or eliminated. Alternately, if $G^*(g)$ is a monotonically decreasing function of g , then $G^*(g(x, y))$ can be used in place of $g(x, y)$ with similar results. The table on the following page summarizes the effect of various functions G and G^* .

$G(H(x, y))$	$G^*(g(x, y)) = b$	x	$\min G(H(b))$	λ
$x^2 + y^2$	$xy = b$	$(b)^{1/2}$	$2b$	2
$(x^2 + y^2)^{1/2}$	$(xy)^{1/2} = b$	b	$(2)^{1/2}b$	$(2)^{1/2}$
$x^2 + y^2$	$(xy)^{1/2} = b$	b	$2b^2$	$4b$
$(x^2 + y^2)^2$	$xy = b$	$(b)^{1/2}$	$4b^2$	$8b$

The first row shows the effect of using $G(H) = H^2$, while retaining the same constraint function. The minimizing point is still at $x = y = (b)^{1/2}$, producing $\min G(H(b)) = 2b$ or $\min H(b) = (2b)^{1/2}$ as expected. The LM, however, has changed from $1/(2b)^{1/2}$ to 2 , i.e., λ is constant for all values of b ; we now have the "constant multiplier". Since the LM is constant for all values of b , the graph of $\min G(H(b))$ is a straight line with slope 2, i.e., the graph is no longer concave downward, but is a straight line. This is progress in the right direction. If instead, we use $G(H) = H$, but $G^*(g) = g^{1/2}$ as in the second row, the result is similar, a constant multiplier, but with a different value. The last two rows show the effects of combining G and G^* . The resulting functions $\min G(H(b))$ are concave upward, no gap exists, and the direct method would solve the problem easily.

Since this is a single constraint problem involving homogeneous functions, the results of using the functions G and G^* above can be read off directly by inspection, without having to solve the problem first. This would permit the analyst to make an appropriate choice of G and G^* prior to attempting the direct method. Using corollary 1, section VII, "p/q" for each of the examples in the above table is 1, 1, 2, 2 (descending order). These match with the exponents of "b" in the column headed " $\min G(H(b))$ ".

Example 7: This problem illustrates some of the simplifications which may be possible in the direct method if the functions involved are homogeneous.

It is desired to minimize $H(x) = \sum_{i=1}^n a_i x_i$ subject to $g(x) = \sum_{i=1}^n |x_i|^{1/2} = C$.
 S is all of n -space.

Solution: The n -space surface $g(x) = C$ has cusps at $\pm C^2$ on each of the coordinate axes. A little reflection about the shape of this surface will make it clear that the minimizing point(s) must be one (or more) of these cusps, at which $g(x)$ is not differentiable. Hence, the classical method cannot be expected to work. If n is small, it may be practical to locate the desired cusp by hand calculations, but if n is large, and several values of C must be examined, computer assistance will be welcome.

The functions $H(x)$, $g(x)$ are homogeneous of degree 1, $1/2$, respectively, hence $p/q = 2$, and the direct method will work. The function $\min H(C)$ is of form AC^2 , so that any guess greater than zero for the LM should produce a meaningful result.

After making a guess for the LM ($\lambda = 1$, say), a computer may easily be programmed to evaluate $F(x)$ at each of the cusps, determine the minimizing cusp, and print out the resulting C' and $\min H(C')$. These values are plugged into the equation

$$\min H(C) = AC^2$$

giving the result that $A = \min H(C')/C'^2$. This solves the problem for all values of C .

Example 8: This problem illustrates the direct method in which the LM remains as a parameter throughout the minimization.

Find the minimum of $H(x, y) = x^2 + y^2$ subject to the constraint $g(x, y) = y - (x-1)^{3/2} = 2$.

Step 1: $F(x, y) = x^2 + y^2 - \lambda y + \lambda(x-1)^{3/2}$

Step 2: $F(x, y)$ can be minimized with respect to x by inspection, obtaining $x = 1$ as the minimizing value (note that x cannot be less than 1 because of the exponent). $F(x, y)$ reduces therefore to $y^2 - \lambda y + 1$ which is parabolic in y . This function has a minimum at $y = \lambda/2$.

Step 3: Hence $g(x, y)$ becomes $\lambda/2$. By the main theorem, section I, we know that the point $x' = (1, \lambda/2)$ provides an absolute minimum for $H(x, y)$. By setting $\lambda = 4$, the required constraint $g(x, y) = 2$ is satisfied.

Although it is not obvious by inspection of the problem, the classical method will fail to solve this problem since the partial derivative of $F(x, y)$, with respect to x , does not exist at the minimizing point. The constraint function $g(x, y)$ is in fact discontinuous at the minimizing point.

There is no need to guess a value for the LM in step 2 since the minimization done there holds for all positive values of the LM (at least) rather than a single value. Normally, when discontinuous functions are involved, the minimization must be done numerically, which requires that the LM be assigned some definite value.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified

1. ORIGINATING ACTIVITY (Corporate author) Operations Evaluation Group, Center for Naval Analyses of the Franklin Institute		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP None	
3. REPORT TITLE The Lagrange Multiplier — A Heuristic Presentation			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Research Contribution			
5. AUTHOR(S) (First name, middle initial, last name) Nunn, W.R.			
6. REPORT DATE 31 October 1966		7a. TOTAL NO. OF PAGES 35	7b. NO. OF REFS 2
8a. CONTRACT OR GRANT NO. NONR 3732 (00)		9a. ORIGINATOR'S REPORT NUMBER(S) Operations Evaluation Group Research Contribution No. 76	
b. PROJECT NO. -----			
c. -----		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None	
d. -----			
10. DISTRIBUTION STATEMENT Distribution of this document is unlimited.			
11. SUPPLEMENTARY NOTES None		12. SPONSORING MILITARY ACTIVITY Office of Naval Research Department of the Navy Washington, D.C. 20350	
13. ABSTRACT Many problems in operations research require the maximization or minimization of a suitable payoff function subject to various constraints. Lagrange multipliers are classically used for this type of problem, however, the treatment given this technique by most texts requires that the payoff and constraint functions be at least differentiable at the extremizing point. This paper shows that the Lagrange multiplier concept can be independent of differentiability or even continuity of the functions involved. It also gives the reader a geometric insight into the workings of the multiplier. Simplifications possible, if the functions involved are homogeneous, are displayed. The treatment is heuristic rather than formally rigorous. (Unclassified)			

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Lagrange multiplier maximization-minimization constraints payoff set differentiability homogeneous functions continuity (Unclassified)						